AD_____

Award Number: W81XWH-10-1-1064

TITLE: Validation of Computerized Cognitive Assessment in Cross-Cultural Populations

PRINCIPAL INVESTIGATOR: C. Munro Cullum, Ph.D.
Jay H. Shore, M.D.
Myron Weiner, M.D.

CONTRACTING ORGANIZATION: University of Texas Southwestern Medical Center at Dallas
Dallas, TX 75390

REPORT DATE: April 2013

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
April 2013	Final	27 September 2010 - 26 March 2013

4. TITLE AND SUBTITLE

Validation of Computerized Cognitive Assessment in Cross-Cultural Populations

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-10-1-1064

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
C. Munro Cullum, Ph.D.

E-Mail: munro.cullum@utsouthwestern.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Texas Southwestern Medical Center at Dallas
Dallas, TX 75390

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The ANAM4 has been used as a pre-deployment tool to evaluate cognitive function, but little is known about possible cultural/ethnic effects. This project compared ANAM4 with traditional neuropsychological measures in healthy African American, American Indian, and Hispanic subjects to assess convergent validity and compare groups in order to determine whether there might be a need for adjusted or separate test norms. Results suggest convergent validity and generally similar ANAM4 results across groups, with the exception of slower reaction time scores compared with published norms, and only minor group differences on certain subtests. Preliminary findings support the use of ANAM4 in cross-cultural populations.

15. SUBJECT TERMS ANAM4, Traumatic Brain Injury(TBI), Computerized Neuropsychological Test, Traditional Neuropsychological Tests, Paper-Pencil Tests Construct Validity

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	13	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

Introduction

The Automated Neuropsychological Assessment Metrics - Version 4 (ANAM4) has been used as a pre-deployment tool to help evaluate the cognitive effects of TBI in the military. Whereas there is support for ANAM4's validity, little is known about what constitutes "normal" performance on ANAM4 among cultural/ethnic minority populations, despite their disproportionate representation in the military, the risk of TBI and PTSD during deployment, and the fact that demographic/cultural factors are known to influence performance on many neurocognitive tests. The primary objectives of this project were to compare the ANAM4 battery with traditional neuropsychological measures of attention, processing speed, and working memory in African American, American Indian, and Hispanic subjects to ensure construct validity in these populations, and to gain a better understanding of the performance of African American, American Indian, and Hispanic men and women on the ANAM4 to determine whether there might be a need for separate or adjusted normative data for these groups.

Body

The project successfully met all of the milestones outlined in Table 1.

TABLE 1: STATEMENT OF WORK

Milestone	Original Plan Date	Completion Date	Remark: Reporting Period
Obtain and set up equipment, pilot testing	11-Jan-2011	27-Dec-2010	27-Sept-2010 - 26-Dec-2010
Staff training and database development	11-Jan-2011	27-Dec-2010	27-Sept-2010 - 26-Dec-2010
Recruitment and testing	26-Oct-2012	27-Aug-2012	27-Sept-2010 - 26-Sept-2012
Data entry	26-Oct-2012	3-Oct-2012	27-Sept-2010 - 26-Oct-2012
Data analysis	26-Mar- 2013	15-Apr-2013	27-Sept-2012 - 26-Apr-2013

Recruitment and testing of subjects

- USAMRMC ORPHRPO Approval received 1/13/2011. Recruitment started 1/14/2011.
- Since recruitment of culturally diverse healthy young controls for this study was anticipated to be a challenge, a variety of recruitment strategies were utilized:

  - "Health Topics Survey Registry" at UT Southwestern Medical Center, which was established as a project to evaluate challenges faced in subject recruiting.
  - "Find a Clinical Trial" at UTSW website, which was developed to connect researchers and community members interested in research.
  - "Campus News," a weekly email updating the UTSW community about research opportunities on campus.
  - African American and Hispanic subjects were recruited in the "Neighborsgo" section of *The Dallas Morning News* from 12/9/2011-12/23/2011.
  - Flyers placed at local businesses and colleges in Dallas and surrounding cities.
  - Coordinator attended Dallas County Health Fairs to publicize study.
  - All Choctaw Nation HSC employees were sent emails about the study.
  - Ads placed in "Bishinik," the Choctaw Nation of Oklahoma newspaper.
  - Flyers disseminated at clinics, health fairs, local businesses, and community colleges in southern Oklahoma.

- Oversampling was deemed necessary for quality control purposes, as it was anticipated that data from some subjects would be excluded because of poor effort on ANAM (as is common with computerized neurocognitive tests). A total of 196 subjects completed testing, with data from 185 deemed valid at first pass (i.e., 11 had questionably valid scores and/or obviously poor effort and their scores were clear outliers). Table 2 presents recruitment targets and sample sizes by group.

TABLE 2: Study Recruitment by Race/ethnicity Group

Race/Ethnicity	Target Sample Size	Final Sample
American Indian	60	63
African American	60	60
Hispanic	60	62
Subtotal	180	185

- Initially, we encountered a 45% no-show rate after telephone screening and appointments were set. To address this, payment incentive was increased from a $25 to $50 gift card, which reduced the no-show rate to 10%.
- No adverse events occurred during the course of the study.
- Many participants reported that the study was interesting and enjoyable.

Study Visit

Subjects were screened via telephone and in-person to determine study eligibility. Individuals that passed the screen and were interested in participating were scheduled for a study visit. Directions were mailed or emailed to subjects, and a reminder call was made the day before the appointment. As noted, our no-show rate was reduced from 45% to 10% when we raised the payment incentive. Study visits took place in the Dallas-Fort Worth, TX and Talihina, OK areas and took 60-90 minutes to complete. Before arriving for the study visit, subjects were randomly assigned to first receive the face-to-face tests or the ANAM4 computerized tests. Once both testing sessions were completed, a satisfaction survey was completed by subjects to provide information about whether they had a preference for testing condition. A list of the tests administered is presented in Tables 3a and b, as per the original proposal.

TABLE 3a: Traditional Neuropsychological Tests

Working and Declarative Memory

- Letter-Number Sequencing (WAIS-IV; Wechsler, 2009): A working memory task where subjects repeat scrambled strings of digits and letters in alphanumeric order.
- Hopkins Verbal Learning Test-Revised (HVLT-R; Brandt & Benedict, 2001): A popular word list-learning measure of verbal declarative memory.

Processing Speed

- Coding subtest (WAIS-IV): A processing speed test wherein subjects must rapidly write in the numbers that are associated with specific symbols.
- Trail Making Test (Army Individual Test Battery, 1944): A popular measure of attention, psychomotor speed, and cognitive sequencing. Includes Parts A and B, with the latter requiring cognitive flexibility.

TABLE 3b. ANAM4 Military Battery

Reaction Time

- Simple Reaction Time: Subjects press a key each time a target character is displayed.
- Procedural Reaction Time: Subjects press one button if a "2" or "3" is presented on screen, and another when a "4" or "5" is displayed.

Working Memory

- Matching to Sample: A 4x4 pattern grid is presented for two seconds, followed by two grids. Subjects must indicate which matches the first one seen.

Processing Speed

- Code Substitution: A series of 10 single-digit numbers paired with a different symbol is presented. Subjects must quickly press a key if the pairing is correct.
- Mathematical Processing: A series of simple arithmetic problems is presented and subjects must indicate if the solution is < than or > five.

Data entry

Data for the study were collected, scored, double-scored, and entered into a database. Data entry was completed 9/12/2012; data cleaning using a standard data verification process was completed 10/8/2013.

Data analysis: Preliminary Results

Preliminary analyses for this project included: a) review of our randomization effectiveness, b) examination of demographic data by group to ensure similarities, c) comparison of ANAM4 with standard neurocognitive tests (Aim 1), and d) initial ethnic group comparisons on ANAM (Aim 2). a) Randomization analysis revealed success in terms of recruitment goals and testing group assignment (i.e., computerized vs. traditional neurocognitive testing condition) for the study. b) Demographic data by group are presented in Table 4.

TABLE 4. Demographic Characteristics x Group

Characteristic	African American (N=63)	American Indian (N=60)	Hispanic (N=62)	Total (N=185)	p-value**
Sex, (% female)	20 (32%)	20 (33%)	20 (32%)	60 (32%)	0.982
Age, mean (SD)	34.7 (7.7)	33.2 (8.3)	30.5 (7.9)	32.8 (8.1)	0.012
Education, mean (SD)	13.6 (1.6)	13.3 (1.5)	13.6 (1.7)	13.5 (1.6)	0.456
Handedness, Right (%)	60 (95%)	54 (90%)	54 (87%)*	168 (91%)	

* 2 were mixed handedness

**c^2 for frequency data, one-way ANOVA for continuous measures

Recruitment and final sample size goals were achieved for each group. Gender representation was targeted for 33% female across the sample, and this was achieved (32%). We were able to recruit subjects of similar age (means ranged from 30.5 to 34.7, with and overall mean of 32.8), although age was significantly different across groups (p=0.012), as African Americans were slightly older than Hispanics (p=0.010) using Bonferroni pairwise post hoc comparisons. Groups were also similar in handedness, with a majority being right-handed. Thus, the demographic composition of each group appears quite similar, which was one of the study goals in order to allow a direct comparison of ANAM4 scores across different ethnic samples.

c) To address Aim 1, a comparison of ANAM4 with standard neurocognitive tests of working and declarative memory and processing speed, Pearson correlations between ANAM4 primary Throughput scores and standard scores from the selected traditional neurocognitive tests were calculated (See Table 5).

Results revealed significant correlations between most ANAM4 and standard neurocognitive test scores in the predicted directions, which generally supports our hypothesis that ANAM4 taps shared cognitive domains with standard clinical tools. However, most of the correlations were in the 0.2 to 0.4 range, reflecting statistically significant yet modest relationships between these different tasks. Correlations in this range are consistent with some previous ANAM validation reports in the literature (e.g. Bleiberg et al., 2000 found correlations in the 0.3 to 0.5 range), but are lower than other reports (i.e., correlations around 0.4 to 0.6) using similar standard neuropsychological measures (e.g. Trails B, WAIS-R Coding) in predominantly Caucasian samples (e.g. Kabat et al., 2001).

TABLE 5: Pearson Product Moment Correlations between Primary ANAM Throughput Scores and Traditional Neuropsychological Tests (N=185)

ANAM4 Measure	Statistic	WAIS4 Letter Number Sequencing	Hopkins Verbal Learning Test	WAIS4 Coding	Trail Making Test A	Trail Making Test B
Simple Reaction Time	*r* =	0.16	0.13	0.07	0.10	0.18
	p-value	0.0249	0.0884	0.3250	0.1781	0.0124
Procedural Reaction Time	*r* =	0.31	0.30	0.38	0.23	0.29
	p-value	**<0.0001**	**<0.0001**	**<0.0001**	**0.0017**	**<0.0001**
Code Substitution Learning	*r* =	0.28	0.33	0.41	0.18	0.32
	p-value	**0.0001**	**<0.0001**	**<0.0001**	0.0148	**<0.0001**
Math Processing	*r* =	0.25	0.31	0.43	0.33	0.34
	p-value	**0.0007**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**
Matching to Sample	*r* =	0.24	0.21	0.30	0.14	0.28
	p-value	**0.0008**	0.0039	**<0.0001**	0.0498	**0.0001**
Code Substitution Delayed	*r* =	0.22	0.27	0.20	0.13	0.09
	p-value	0.0024	**0.0002**	0.0077	0.0695	0.2066

Note: Bolded p-values are those significant at the <.001 level using Bonferroni correction

Not unexpectedly, simple reaction time was unrelated (correlations < .20) to any of the standard neuropsychological tests, while Procedural Reaction Time and Math Processing scores showed significant relationships with all five standard neuropsychological tests. Math Processing has generally showed good correlations across many standard neuropsychological tests in the literature (e.g. see Short et al., 2007) and appears to rely upon a variety of cognitive skills beyond simple calculation (e.g. working memory, processing speed). Code Substitution was correlated with all tests except Trail Making A, and Matching to Sample was related to Letter-Number Sequencing, WAIS Coding, and Trail Making B. These findings suggest similar correlations across most tasks examined, with the strongest and most consistent relationships evident between ANAM4 scores and the WAIS4 Coding subtest (mean r = 0.34, excluding Simple Reaction Time). Thus, the hypothesized associations between tests that measure more similar cognitive constructs (i.e., processing speed and working or episodic memory) were not generally supported by the present results, since most measures showed similar correlations across tasks in these samples. Since most neurocognitive measures involve a variety of cognitive abilities beyond their main "focus," however, the numerous nonspecific correlations are not surprising, and it appears that ANAM4 shows general correlations with cognitive functioning in a global, nonspecific fashion in ethnically diverse populations.

d) For Aim 2, which sought to compare racial/ethnic groups on ANAM4, groups were compared using ANOVA. Because the groups were similar across demographic factors (aside from a small but statistically significant difference in age) and because age-adjusted standard scores from the standard neurocognitive tests were used, ANOVA was performed without covariates for the initial analysis. Table 6 presents means, standard deviations, and ANOVA results of the groups and the total sample on the primary ANAM throughput scores.

TABLE 6. Means, Standard Deviations, and One-way ANOVA Results by Race/Ethnicity for ANAM4 Throughput measures

ANAM4 Measure	African American (N=63)		American Indian (N=60)		Hispanic (N=62)		Total (N=185)		ANOVA	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	F 2,182	p-value
Simple Reaction Time	193.2	29.8	192.4	35.9	200.2	34.0	195.3	33.3	1.0	0.3573
Procedural Reaction Time	92.0	16.3	96.8	12.9	96.0	15.6	94.9	15.1	1.9	0.1579
Code Substitution Learning	48.9	12.5	51.3	11.4	52.7	11.2	50.9	11.7	1.72	0.1812
Math Processing	21.8	6.3	21.5	6.2	22.0	5.8	21.7	6.1	0.13	0.8765
Matching to Sample	27.5	9.5	33.7	13.2	35.5	9.9	32.2	11.4	9.0	0.0002
Code Substitution Delayed	34.5	16.6	39.1	15.1	40.7	19.0	38.1	17.1	2.2	0.1129

All ANAM4 Throughput measures were similar across groups except the Matching to Sample score ($p=0.0002$), wherein African Americans performed significantly lower than American Indians ($p=0.006$) and Hispanics ($p<0.001$) using Bonferroni pairwise post hoc comparisons. This suggests that most ANAM4 subtests appear to be minimally influenced by ethnicity in these samples, supporting the original hypothesis.

Underscoring this interpretation is the overall similarity between the current results from all three groups considered together in reference to previously published military norms of similar-age subjects (Vincent et al., 2012). To illustrate, means and standard deviations of ANAM4 Throughput scores from the current total sample are presented in Table 7 along with representative values (based on age 31-35) from a large ($N > 12,000$), primarily Caucasian active-duty military normative sample (Vincent et al., 2012). Simple Reaction Time scores were lower in the current overall sample by just over one standard deviation compared to published normative age-reference values, but all other mean scores on ANAM4 tests were highly similar across groups (See Table 7).

TABLE 7. ANAM4 Means and Standard Deviations: Total Sample vs. Published Norms

ANAM4 Subtest	Current Sample (N=185)	Vincent et al. (2012)
Simple Reaction Time	195.3 (33.3)	233.2 (30.8)
Procedural Reaction Time	94.9 (15.0)	99.2 (14.2)
Code Substitution Learning	50.9 (11.7)	50.6 (10.6)
Math Processing	21.7 (6.1)	22.4 (6.4)
Matching to Sample	32.2 (11.4)	33.4 (10.0)
Code Substitution Delayed	38.1 (17.1)	42.0 (15.0)

It is noteworthy that in addition to similar mean scores across these different samples, the standard deviations were quite consistent, suggesting similar psychometric test characteristics among healthy young adults in these cross-cultural groups. Further analysis of groups by gender will be needed to explore potential interaction effects, and additional exploration of the lower Simple Reaction Time scores in the present samples versus published norms is also in order, along with careful examination of the lower scores of the African American group on the Matching to Sample subtest. Whether adjusted norms might be needed for simple reaction time is a question, but the more cognitively demanding tasks appear to work well in cross-cultural groups similar to these.

Key Research Accomplishments

Recruitment of all subject group target numbers was successful. As with any computer-based neurocognitive testing endeavor, a small proportion will be invalid due to inadequate effort, lack of adequate reading test instructions, and other factors. We excluded results from 11 subjects as a result of their scores being deemed invalid as significant outliers. Randomization efforts were effective in terms of the order of test condition (i.e., ANAM4 vs. standard tests), and we were successful in obtaining well-matched samples in terms of education and gender representation, which were key goals in allowing for comparison of ANAM results across groups. All groups were similar in age, although the difference between African American and Hispanic samples (34.7 vs. 30.5) was statistically significant ($p = .01$). This difference is considered negligible from a clinical/developmental standpoint, however, and as a result, preliminary group comparisons were done using ANOVA as noted above, without adjustment for age (particularly since age-adjusted neurocognitive test scores were used). As such, the groups appear very well-balanced from a demographic perspective. Importantly, our overall results appear generally quite consistent with previously published military norms.

Data analysis has been completed for the primary aims as noted above. Preliminary reportable outcomes are described below.

Reportable Outcomes

Initial findings suggest generally significant (though modest) correlations between results from ANAM4 and standard neurocognitive measures of processing speed and working and episodic memory, providing overall support for our hypothesis of ANAM4's convergent validity. However, the magnitude of the correlations was modest and ANAM4 showed similar correlations across the standard neuropsychological tests, regardless of primary cognitive domain. As such, ANAM4 seems to represent a general "cognitive" factor based upon these analyses in these populations. In terms of the comparison of different ethnic groups on ANAM4, general similarities in test performance across Native American, African American, and Hispanic subject groups were seen across the present samples. Slower reaction times were seen in all three of these groups compared with Caucasian norms, and African Americans obtained slightly lower scores on the Matching to Sample subtest of ANAM4 than the American Indian and Hispanic samples, but no other differences were suggested in preliminary analyses. This suggests that ANAM4 can be used in non-Caucasian populations and that when education, gender, and (to a lesser extent) age are similar, ethnic differences in ANAM4 performance appear minimal, with the possible exception of simple reaction time scores, which merits further investigation. Whereas additional analyses need to be completed before firm conclusions can be drawn, these initial findings suggest good convergent validity and clinical utility of most ANAM4 subtests in African American, Native American, and Hispanic individuals. As such, results appear to support the use of ANAM4 in these populations without the need for specialized ethnic norms, which has long been a question in the use of ANAM4.

We are in the process of continuing primary data analysis to verify the above-noted findings and prepare the primary manuscript summarizing results. As part of this process, we will explore potential covariates that are of interest (e.g. details regarding educational history), in addition to comparing the groups by gender, and further examining relationships between select ANAM4 and traditional neurocognitive test scores. Understanding the slower simple reaction time scores in the current samples compared with published norms is also in order.

Conclusion

Preliminary data analyses suggest convergent validity of ANAM4 in African American, Native American, and Hispanic samples in terms of the correlations between ANAM4 and standard neurocognitive tests that tap similar cognitive domains. Furthermore, generally similar ANAM4 results were obtained across the different ethnic groups examined, and mean overall results are similar to published military norms in similar age samples. These findings are promising and lend further support to the clinical utility of ANAM4 in ethnically diverse populations without the need for separate racial/ethnic norms for interpreting scores. As such, the primary aims of the current study were achieved, and hypotheses supported.

Plans for the future include further summarization and preparation of the above-noted findings for publication in peer-reviewed journals. This will include several manuscripts as various other ANAM4 scores are analyzed (e.g. % correct x subtest), in addition to examination of the effects of gender and more detailed analysis of education across measures. Breakdowns of scores by percentiles and further comparison with existing military and civilian norms will also be prepared, in addition to exploration of more detailed scores and subscores from the various standard neuropsychological tests that were administered. We will also

analyze the data we have collected regarding consumer satisfaction with computerized versus traditional cognitive testing.

References

Kabat, M.H., Kane, R.L., Jefferson, A.L., & DiPino, R.K. (2001). Construct validity of selected Automated Neuropsychological Assessment Metrics (ANAM) Battery measures. *The Clinical Neuropsychologist*, 15, 498-507.

Short, P., Cernich, A., Wilken, J.A., & Kane, R.L. (2007). Initial construct validation of frequently employed ANAM measures through structural equation modeling. *Archives of Clinical Neuropsychology*, 22S, S63-S77.

Vincent A.S., Roebuck-Spencer T, Gilliland K, Schlegel R. (2012). Automated Neuropsychological Assessment Metrics (v4) Traumatic Brain Injury Battery: Military Normative Data. *Military Medicine, 177*, 3:256-269.

Appendices

N/A

Title: **Validation of Computerized Cognitive Assessment in Cross-Cultural Populations**

Proposal ID: 09204005 Funding Source: USAMRMC

PI: C. Munro Cullum, Ph.D. Org: University of Texas Southwestern Medical Center Award Amount: $509,066

DMRDP

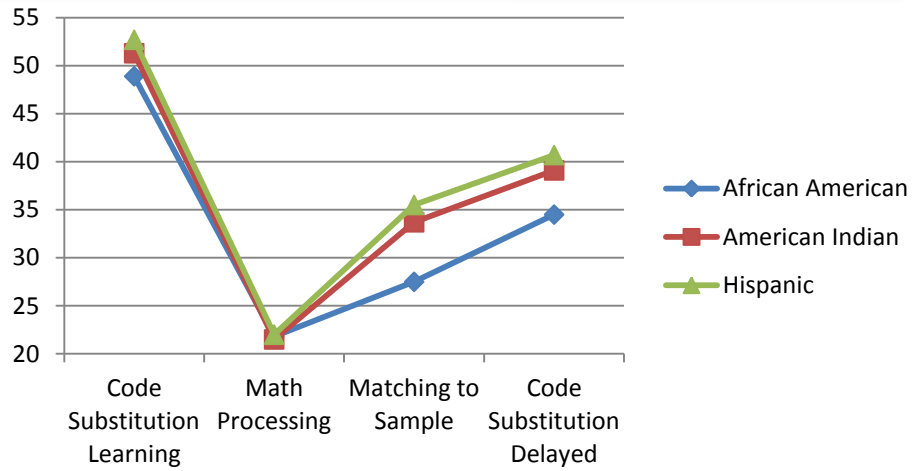
Study Aims

A) Compare the ANAM4 computerized test battery with traditional cognitive tests to show construct validity in cross-cultural populations.

B) Compare performance of African American, American Indian, and Hispanic groups on the ANAM4 and determine whether separate or adjusted normative data based on race/ethnicity are needed.

Approach

- Recruit and test groups of healthy African American, American Indian, and Hispanic men and women on ANAM4 and standard neuropsychological tests.
- 185 valid test sessions completed including ANAM4 and traditional tests of working memory, episodic memory, and processing speed.
- Statistical comparison of test results across ethnic/cultural groups using correlational analysis and ANOVA.



Despite slower reaction time scores compared with published norms, similar performances were seen on most ANAM4 scores across groups, supporting utility and robustness of ANAM4 in cross-cultural populations.

Time	Milestone	Cost
Grant Year 1	•Obtain and set up equipment, pilot testing •Staff training and set up database •Recruitment and testing in progress •Data entry in progress	•Budgeted $225,402 •Spent $188,415
Grant Year 2	•Recruitment and testing of well-matched subjects x cultural/ethnic group •Data entry & verification completed	•Budgeted $283,664 •Spent $240,277
Six month no cost extension (NCE)	•Preliminary data analysis completed	•Spent remaining $80,374

Milestones

GY1

✓Obtained, set up, and tested equipment, hired staff, created database

✓Recruitment procedures established, testing subjects started.

GY2

✓Recruitment and testing of target samples completed (N= 185)

✓Data entry & verification completed.

NCE

✓Initial data analysis completed. Data summaries underway.

Key Accomplishments

- Recruitment of target numbers of well matched samples.
- Initial findings suggest: a) significant but modest correlations between ANAM4 and standard cognitive tests, and b) nominal effects of cultural background on most ANAM4 results with the exception of reaction time scores, which were lower across groups than published norms.

Projected Expenditure:$509,066 Actual Expenditure: $509,066